Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene

Daniel R. Larson, *et al.*

Science **332**, 475 (2011);
DOI: 10.1126/science.1202142

Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene

Daniel R. Larson,[1,2,3]* Daniel Zenklusen,[1,4] Bin Wu,[1,2] Jeffrey A. Chao,[1] Robert H. Singer[1,2,3]†

Cellular messenger RNA levels are achieved by the combinatorial complexity of factors controlling transcription, yet the small number of molecules involved in these pathways fluctuates stochastically. It has not yet been experimentally possible to observe the activity of single polymerases on an endogenous gene to elucidate how these events occur in vivo. Here, we describe a method of fluctuation analysis of fluorescently labeled RNA to measure dynamics of nascent RNA—including initiation, elongation, and termination—at an active yeast locus. We find no transcriptional memory between initiation events, and elongation speed can vary by threefold throughout the cell cycle. By measuring the abundance and intranuclear mobility of an upstream transcription factor, we observe that the gene firing rate is directly determined by trans-activating factor search times.

Transcription initiation occurs through a series of sequential steps initiated by the binding of gene-specific activators and coactivators and resulting in recruitment of the basal transcription machinery and RNA polymerase (*1*). Genome-wide chromatin immunoprecipitation (ChIP) experiments have provided detailed information about the occupancy of transcription factors on DNA, and live-cell photobleaching recovery studies on tandem gene repeats indicate that turnover of these factors is generally on the order of seconds (*2*–*4*). These approaches describe the upstream regulators of transcription in a cellular context. However, to gain a functional understanding of transcription networks, it is necessary to observe the output of transcription, namely RNA production. We have developed a single-molecule assay to observe the enzymatic activity of an RNA polymerase II (RNAPII) molecule directly on an active gene in a living cell. This methodology enables direct measurement of promoter clearance, elongation, and termination rates. In *Saccharomyces cerevisiae*, we find that elongation is deterministic and initiation is stochastic. Moreover, by comparing nascent RNA dynamics with the reaction-diffusion behavior of an upstream transcription factor, we determine that promoter firing rate is dictated directly by the search times in the nucleus required for the factor to find the gene.

To detect RNAPII activity in living cells, we used a green fluorescent protein (GFP)–labeling approach adapted from the MS2 technique (*5*, *6*). A cassette coding for 24 binding sites for the PP7 bacteriophage coat protein was inserted into the 5′ untranslated region (5′UTR) of a target gene (*7*). A PP7-GFP fusion protein is constitutively

[1]Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [2]Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [3]Transcription Imaging Consortium, Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA. [4]Département de biochimie, Université de Montréal, Montréal, Québec H3T 1J4, Canada.

*Present address: Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA.
†To whom correspondence should be addressed. E-mail: robert.singer@einstein.yu.edu

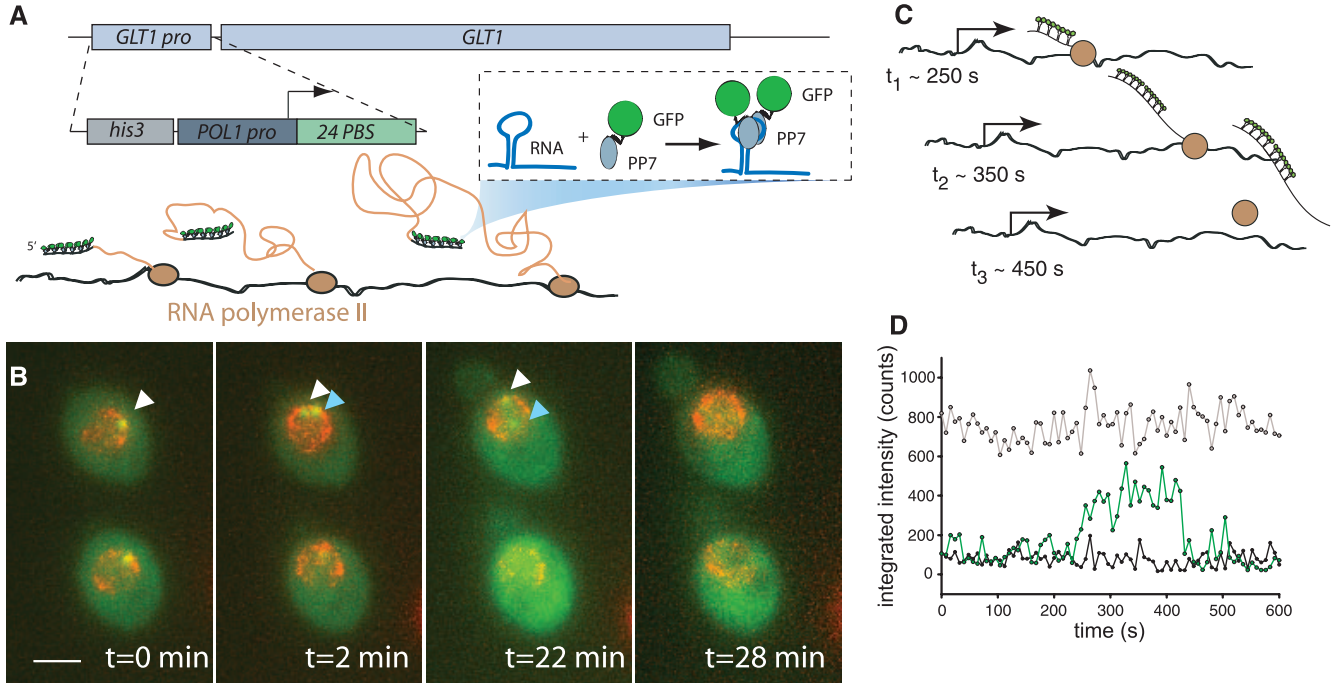

Fig. 1. Real-time measurement of fluorescent RNA reveals transcription kinetics. (**A**) Diagram of the *POL1pro-GLT1* reporter gene, obtained by replacing the *GLT1* promoter with a cassette containing a his5 marker, *POL1* promoter, and 24 PBS. (**B**) Time-dependent activity of individual reporter genes (PP7-GFP, green; Nup49-tomato, red). At time t = 0 min, both cells show a TS near the periphery of the nucleus. At t = 2 min, an additional TS corresponding to the duplicated gene has turned on in the upper cell (white and blue arrows). At t = 22 min, the nascent bud is visible in both cells, and the TSs of the upper cell are present at opposite sides of the nucleus, but the lower cell TS has turned off. At t = 28 min, the bud continues to grow, and both TSs have turned off. All images are maximum projected z-stacks. Scale bar indicates 3 μm. (**C** and **D**) Diagram of RNAPII progression on the gene with corresponding intensity trace (green line). The gray line is the intensity of a cytosolic mRNA; the black line is a background intensity at an arbitrary position in the nucleus measured by using the fitting algorithm (movie S2).

co-expressed and binds the stem loops after they have been transcribed, resulting in a GFP-labeled transcript (Fig. 1A) (*8*–*10*). Thus, one can directly measure both the earliest steps in pre-mRNA synthesis and the total dwell time of RNA at the site of transcription. The cell cycle–regulated *POL1* promoter (*POL1* pro) and PP7 binding site cassette (24 PBS) were knocked in upstream of the endogenous *GLT1* gene, which was then found to transcribe similarly to the endogenous *POL1* gene (fig. S1). When monitored by using wide-field microscopy, nascent transcription sites (TSs) are detected as diffraction-limited fluorescent puncta in the nucleus, the boundary of which is determined by using a red nuclear pore marker (Fig. 1B). The *POL1* promoter was active during late G_1/S phase, but activity periods varied from cell to cell and even for different alleles within the same nucleus, as shown in a time series in two different cells (Fig. 1B and movie S1).

To analyze transcription kinetics, we followed the TS in the nucleus over time. The position and integrated intensity of an individual TS was determined by using a tracking algorithm (*11*) to generate the spatial trajectory and intensity time trace of the site (Fig. 1, C and D). Three distinct phases in the time trace are evident: (i) the increase in fluorescence signal corresponding to polymerization through the PP7 cassette as PP7-GFP binds recently transcribed RNA stem loops, (ii) a stable intensity of fluorescence corresponding to transcription of the *GLT1* gene, and (iii) a sharp drop in fluorescence that corresponds to termination and release of the nascent RNA, which then rapidly diffuses away from the site (movies S2 and S3).

For an actively transcribing gene, many such events will be superimposed, with RNAPIIs in various positions along the gene, necessitating an analysis approach that takes into account these nonequilibrium kinetic processes (Fig. 2, A and B, and movie S4). To extract kinetic parameters from these transcription time traces, we developed an analysis approach that takes advantage of the fluctuations inherent in the measurement (*12*, *13*). An idealized fluorescence time trace for transcription of a single pre-mRNA is shown in Fig. 2C. This process can be described by an autocorrelation function $G(\tau)$, which is a discrete autocorrelation over all the transition probabilities for RNAPII (*6*). The variables are the total dwell time of a transcript (T), which includes elongation and termination, and the transcript initiation rate (c). For long genes, the solution takes the form

$$G(\tau) = \frac{(T-\tau)}{cT^2} H(T-\tau) \qquad (1)$$

where τ is the autocorrelation delay and H denotes the Heaviside step function, which is unity when $\tau \leq T$ and zero when $\tau > T$ (Fig. 2D). Therefore, by steady-state observations of TS fluctuations it is possible to measure both the transcript initiation rate and the total time a transcript spends at the TS.

The autocorrelation for the *POL1p-GLT1* construct reflects deterministic elongation times and no correlation between individual transcripts (Fig. 2E). The autocorrelation shows a linear decay with a sharp transition at the time that corresponds to the dwell time of a transcript at the TS ($T = 290 \pm 30$ s) (SEM). The processive nature of pre-mRNA synthesis, consisting of thousands of sequential enzymatic cycles, results in a central-limit process where the total dwell time is distributed around a nonzero mean. Alternatively, a comparable number of micropauses interspersed

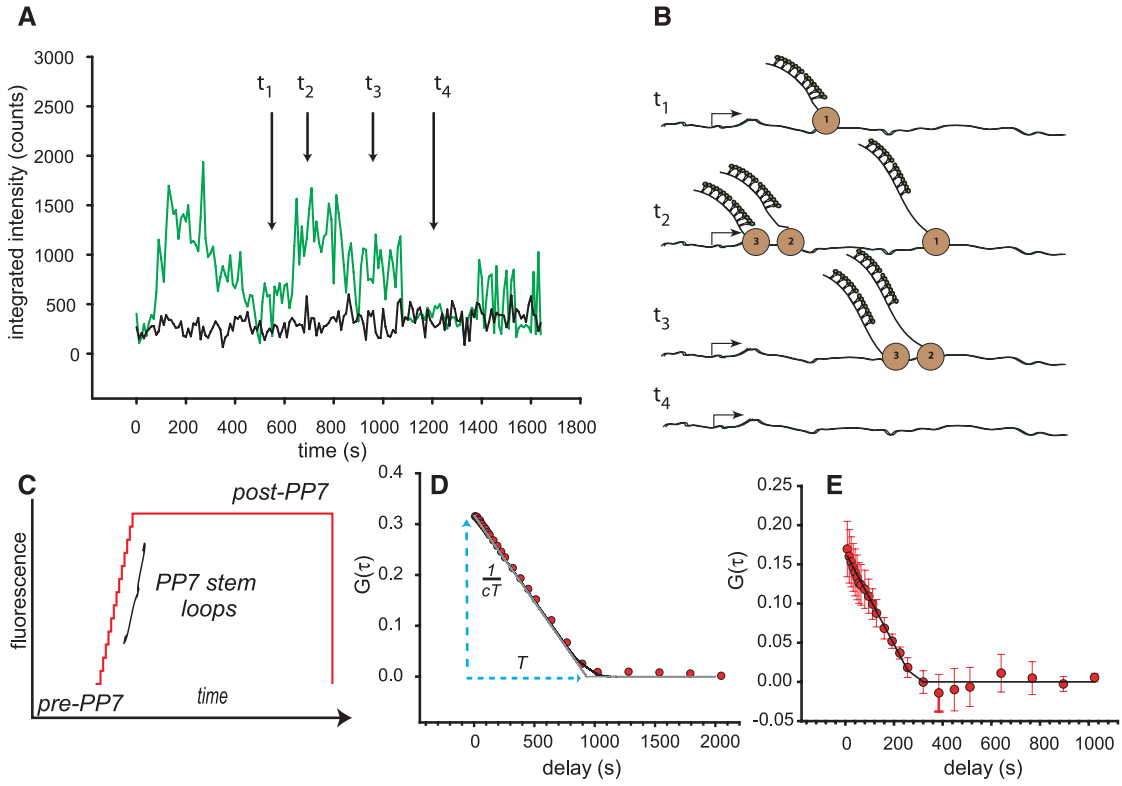

Fig. 2. Direct measurement of transcription initiation and elongation of the *POL1p-GLT1* gene with fluctuation analysis. (**A**) TS intensity trajectory (green) compared to background (black). (**B**) Diagram of RNAPII occupancy based on the intensity data from (A). At time t_1, a single RNAPII is loaded on the gene; at t_2, multiple RNAPIIs load in rapid succession before the first has finished polymerization; at t_3, the first RNAPII has terminated; at t_4, the TS is devoid of nascent RNA. (**C**) Diagram of a single transcript intensity trajectory. After promoter clearance, RNAPII proceeds through the 24 stem loops, resulting in a discrete fluorescence increase upon PP7-GFP binding. After this 1.5-kb cassette, RNAPII continues through the gene, and the fluorescence is level. Upon completion of the transcript, there is a decrease corresponding to the brightness of a single transcript. (**D**) Autocorrelation function for transcription. The analytical solution (black curve) is compared to a Monte Carlo simulation (red circles) and the approximation for a long gene (Eq. 1) (gray curve). (**E**) Autocorrelation for *POL1p-GLT1*. The autocorrelation (red circles) is fit with Eq. 1. n = 10 cells. Error bars indicate SEM.

with polymerization cycles would also produce a similar autocorrelation (*14*) (fig. S2). In addition, the autocorrelation vanishes at times greater than the dwell time, indicating that there is no transcriptional memory between initiation events. Thus, although the *POL1* promoter is only active during a certain stage of the cell cycle, initiation events during this active period are stochastic and uncorrelated and occur with a frequency $c = 1.3 \pm 0.72$ min^{-1}.

We then sought to separate the kinetic processes of initiation, elongation, and termination. By placing the PP7 cassette in either the 5′UTR or 3′UTR of a long gene such as *MDN1* (15 kb), one can measure dwell times in which termination plays a lesser or greater role in comparison to elongation (Fig. 3, A to C) (*15*, *16*). With the cassette in the 5′UTR, the dwell time is dominated by the downstream portion of the gene (*PP7-MDN1*, Fig. 3D, red curve, and movies S5 andS6), where the fluorescence is not changing, leading to the autocorrelation in Eq. 1. With the cassette in the 3′UTR (*MDN1-PP7*, Fig. 3D, blue curve, and movie S7), the autocorrelation reflects the transitions between N stem loops, which occur with frequency k:

$$G(\tau) = \frac{k}{c}\left(\frac{2}{3}\right)\frac{1}{[N(N+1)]^2}e^{-k\tau}\sum_{i=0}^{N}[(N-i) \times (N-i+1)(2N+i+1)\frac{(k\tau)^i}{i!}] \quad (2)$$

In this case, the total dwell time T equals N/k. By using this complementary approach to selectively emphasize early and late events in the lifetime of a nascent RNA, we can separate elongation from termination.

MDN1 transcription dwell times and initiation rates varied over the cell cycle. The *MDN1* gene was monitored at various stages of the cell cycle, resulting in fluorescence intensity traces during G_1 for *PP7-MDN1* and *MDN1-PP7* (Fig. 3, E and F). During G_1, the mean dwell time of *PP7-MDN1* transcripts is $T = 770 \pm 260$ s, and the mean dwell time of *MDN1-PP7* is $T = 140 \pm 30$ s (Fig. 3, G and H). From these data, it is possible to directly compute the elongation velocity according to the equation $v = \frac{(L_{5'} - L_{3'})}{(T_{5'} - T_{3'})}$, where L is the length of the PP7 cassette plus the downstream sequence and T is the dwell time as measured above. The velocity of RNAPII on *MDN1* is $v = 20 \pm 8$ bases per second, and the termination time is 70 ± 41 s. The initiation rates are $c = 0.16 \pm 0.07$ min^{-1} for *PP7-MDN1* and $c = 0.24 \pm 0.1$ min^{-1} for *MDN1-PP7*. The agreement between initiation rates follows from the processivity of RNAPIIs, which clear the promoter and proceed to termination. In late S/G_2 phase, *PP7-MDN1* was observed to have a much shorter mean dwell time ($T = 310 \pm 42$ s) and a faster mean initiation rate ($c = 0.40 \pm 0.18$ min^{-1}) (Fig. 3I and fig. S3), resulting in a lower bound estimate of the velocity of 46 ± 6.2 bases per second. This variable dwell time of single RNAPIIs on the *MDN1* gene is obscured by population averages of the single-molecule data and can be misinterpreted as RNAPII pausing (figs. S4 and S5).

Because *POL1p-GLT1* and *MDN1* each exhibit uncorrelated initiation events, gene firing is likely determined by a single rate-limiting step. Such a step might be the binding of a trans-acting factor to DNA, so we used two-photon fluorescence correlation spectroscopy (2PE-FCS) to determine the abundance and reaction-diffusion properties of the transcription factor Mbp1p (*17*, *18*), which binds the *POL1* promoter in vivo (*2*) and is necessary and sufficient for correct cell cycle–regulated expression of *POL1* (*19*). The 2PE-FCS autocorrelation of Mbp1-GFP under endogenous regulation is shown along with a fit to a model where Mbp1-GFP alternates between bound and unbound states (Fig. 4 and fig. S6) (*20*). The time in the unbound state corresponds to three-

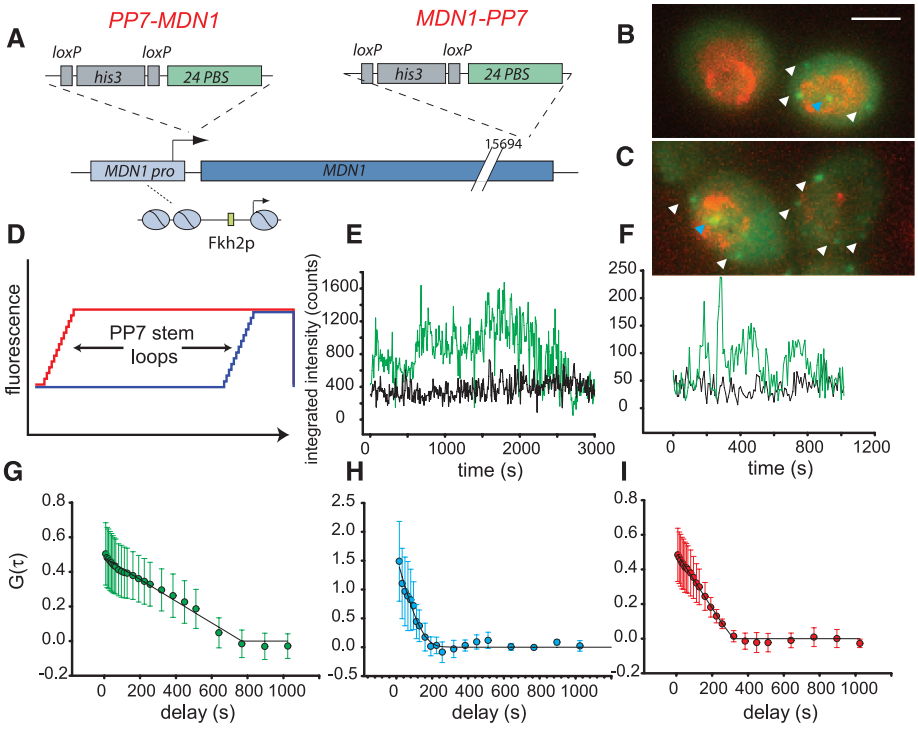

Fig. 3. The housekeeping gene *MDN1* shows variable elongation rate. (**A**) Diagram of the *MDN1* gene with the PP7 cassette inserted either in the 5′ or the 3′UTR of *MDN1* at the endogenous locus. (**B** and **C**) Visualization of TS and cytosolic mRNA (PP7-GFP, green; Nup49-tomato, red) for the *PP7-MDN1* and *MDN1-PP7* genes, respectively. A TS is designated with a blue arrowhead; cytosolic transcripts are designated by white arrowheads. Scale bar, 3 μm. (**D**) Diagram of the fluorescence time traces for *MDN1* with stem loops in the 5′ or 3′UTR, *PP7-MDN1* and *MDN1-PP7*. (**E** and **F**) Intensity trajectories of single *PP7-MDN1* and *MDN1-PP7* sites. (**G**) *PP7-MDN1* transcription kinetics in G_1 cells. The *PP7-MDN1* autocorrelation was fit with Eq. 1. $n = 10$ cells. (**H**) *MDN1-PP7* transcription kinetics in G_1 cells. The *MDN1-PP7* autocorrelation was fit with Eq. 2. $n = 10$ cells. (**I**) *PP7-MDN1* transcription kinetics in late $S/G_2/M$. The *PP7-MDN1* autocorrelation was fit with Eq. 1. $n = 10$ cells. Error bars indicate SEM. Cell cycle stage was determined by size, nuclear morphology, budding index, and presence of multiple alleles (fig. S8).

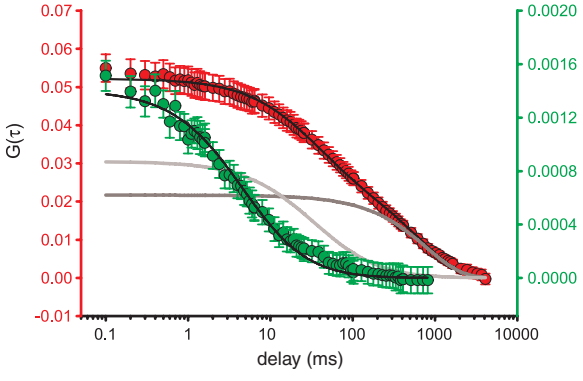

Fig. 4. Mbp1p-GFP transactivator dynamics in the yeast nucleus. 2PE-FCS autocorrelation of Mbp1p-GFP (red circles) and NLS-GFP (green circles), $n = 25$ and 17 cells, respectively. Mbp1p-GFP autocorrelation was fit to a model of reaction diffusion on the basis of transitions between free diffusion (light-gray curve) and nonspecific binding to chromatin (dark gray curve). Concentration of *Mbp1-GFP* = 60 ± 16 nM; $D = 0.6 \pm 0.1$ μm^2/s; $\tau_{3D} = 1.1 \pm 0.2$ s; $\tau_{1D} = 0.8 \pm 0.1$ s. NLS-GFP was fit to a simple diffusion model with $D = 2.7 \pm 0.4$ μm^2/s. Error bars indicate SEM.

dimensional (3D) diffusion within the nucleus, and the time in the bound state corresponds to 1D diffusion along DNA while the factor is nonspecifically bound (*21*). GFP targeted to the nucleus with a nuclear localization sequence (NLS-GFP) (*18*) shows simple 3D diffusion. The number of Mbp1-GFP molecules in the nucleus is 350 ± 18.

From these data, it is possible to directly compute the search time (t_s), according to the Berg and von Hippel model of facilitated diffusion (*22*):

$$t_s = \frac{M}{N\sqrt{4D_{1D}\tau_{1D}}}(\tau_{3D} + \tau_{1D}) \qquad (3)$$

where M is the number of base pairs in the nucleus (12×10^6), D_{1D} (5.1×10^5 bp^2/s) is the 1D diffusion coefficient measured in vitro (*21*), and N is the number of factors doing the search. The calculated search time is 52 ± 8 s, which compared to the measured time between initiation events of 46 ± 25 s (Fig. 2E). A modified description of facilitated diffusion that postulates that the search regions are determined by the nucleosome-free DNA results in a search time of 84 ± 11 s (fig. S7) (*23*). Both models of facilitated diffusion predict a search time that is in quantitative agreement with the direct measurement of transcription initiation frequency for the *POL1* promoter. Thus, the initiation rate from this active state shows a rate-limiting step that is consistent with the search time for Mbp1p, suggesting that every encounter leads to a promoter clearance event.

The view of transcription in yeast that emerges from these data is consistent with a model where transcription activation is determined by recruitment of a rate-limiting factor, which then assembles the pre-initiation complex (fig. S7) (*24*, *25*). Remarkably for the *POL1* promoter, this recruitment process is highly efficient, but the Mbp1p-DNA complex is also dynamic, as is the basal transcription machinery (*26*), resulting in single uncorrelated transcription events instead of the bursts of transcription observed for other genes (*8*, *16*, *27*, *28*). Moreover, transcription rates in yeast are consistent with a search process occurring in a functionally homogeneous nuclear compartment. The single-factor, single-target search time for Mbp1p in the yeast nucleus is ~5 hours. In contrast, the Lac repressor search time in bacteria is ~300 s (*21*). However, the effective search times are similar, because of the number of factors doing the search (for LacI, 20 molecules; for Mbp1, 350 molecules). It is unclear whether the biophysical aspects of the search process are modified in the metazoan nucleus, where many genes are not expressed and chromatin shows substantial compartmentalization. The live-cell approach described in this report, which uses fluctuation analysis to observe the activity of RNAPII and the reaction-diffusion behavior of transcription factors, will be critical in deciphering transcriptional control at single genes operating in their native context.

References and Notes

1. B. J. Venters, B. F. Pugh, *Genome Res.* **19**, 360 (2009).
2. C. T. Harbison *et al.*, *Nature* **431**, 99 (2004).
3. G. L. Hager, J. G. McNally, T. Misteli, *Mol. Cell* **35**, 741 (2009).
4. X. Darzacq *et al.*, *Annu Rev Biophys* **38**, 173 (2009).
5. E. Bertrand *et al.*, *Mol. Cell* **2**, 437 (1998).
6. Materials and methods are available as supporting material on *Science* Online.
7. J. A. Chao, Y. Patskovsky, S. C. Almo, R. H. Singer, *Nat. Struct. Mol. Biol.* **15**, 103 (2008).
8. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, *Cell* **123**, 1025 (2005).
9. X. Darzacq *et al.*, *Nat. Struct. Mol. Biol.* **14**, 796 (2007).
10. S. M. Janicki *et al.*, *Cell* **116**, 683 (2004).
11. D. R. Larson, M. C. Johnson, W. W. Webb, V. M. Vogt, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15453 (2005).
12. P. Cluzel, M. Surette, S. Leibler, *Science* **287**, 1652 (2000).
13. E. Elson, D. Magde, *Biopolymers* **13**, 1 (1974).
14. E. A. Galburt *et al.*, *Nature* **446**, 820 (2007).
15. S. Boireau *et al.*, *J. Cell Biol.* **179**, 291 (2007).
16. D. Zenklusen, D. R. Larson, R. H. Singer, *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).
17. J. Yao, K. M. Munson, W. W. Webb, J. T. Lis, *Nature* **442**, 1050 (2006).
18. B. D. Slaughter, J. W. Schwartz, R. Li, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20320 (2007).
19. A. Pizzagalli *et al.*, *Curr. Genet.* **21**, 183 (1992).
20. A. Michelman-Ribeiro *et al.*, *Biophys. J.* **97**, 337 (2009).
21. J. Elf, G.-W. Li, X. S. Xie, *Science* **316**, 1191 (2007).
22. O. G. Berg, R. B. Winter, P. H. von Hippel, *Biochemistry* **20**, 6929 (1981).
23. L. Mirny, M. Slutsky, Z. Wunderlich, J. L. Tafvizi, A. Kosmrlj, *J. Phys. A* **42**, 434013 (2009).
24. M. Ptashne, A. Gann, *Nature* **386**, 569 (1997).
25. M. Keaveney, K. Struhl, *Mol. Cell* **1**, 917 (1998).
26. R. O. Sprouse *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13304 (2008).
27. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, *PLoS Biol.* **4**, e309 (2006).
28. J. R. Chubb, T. Trcek, S. M. Shenoy, R. H. Singer, *Curr. Biol.* **16**, 1018 (2006).

Acknowledgments: The authors thank the following for critical reading of the manuscript: D. Grünwald, T. Lionnet, T. Trček, J. Yao, X. Darzacq, R. Tjian, and M. Larson. D.R.L. thanks W. Zipfel, S. Hess, J. Atilgan, and M. Nicholas for discussions about fluctuation analysis. S. Silverman provided the fkh1fkh2 deletion strain. B. Slaughter provided the NLS-GFP strain and J. Vogel the Mbp-GFP strain. We thank E. Gruss Lipper for her gift founding the Gruss Lipper Biophotonics Center that provided the equipment used in this study. Supported by NIH National Institute of General Medical Sciences grants 57071 and 86217 to R.H.S. There is a patent (no. 6,586,240) associated with the methods for visualization of RNA in living cells as presented in this manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/332/6028/475/DC1
Materials and Methods
Figs. S1 to S8
References
Movies S1 to S7

23 December 2010; accepted 25 February 2011
10.1126/science.1202142

The Growth Factor Progranulin Binds to TNF Receptors and Is Therapeutic Against Inflammatory Arthritis in Mice

Wei Tang,[1,2]* Yi Lu,[1,2]* Qing-Yun Tian,[1]* Yan Zhang,[1] Feng-Jin Guo,[1]† Guang-Yi Liu,[1] Nabeel Muzaffar Syed,[1] Yongjie Lai,[1] Edward Alan Lin,[1] Li Kong,[1] Jeffrey Su,[3] Fangfang Yin,[4]‡ Ai-Hao Ding,[4] Alexandra Zanin-Zhorov,[5] Michael L. Dustin,[5] Jian Tao,[6] Joseph Craft,[6] Zhinan Yin,[7] Jian Q. Feng,[8] Steven B. Abramson,[9] Xiu-Ping Yu,[2] Chuan-ju Liu[1,10]§

The growth factor progranulin (PGRN) has been implicated in embryonic development, tissue repair, tumorigenesis, and inflammation, but its receptors remain unidentified. We report that PGRN bound directly to tumor necrosis factor receptors (TNFRs) and disturbed the TNFα-TNFR interaction. PGRN-deficient mice were susceptible to collagen-induced arthritis, and administration of PGRN reversed inflammatory arthritis. Atsttrin, an engineered protein composed of three PGRN fragments, exhibited selective TNFR binding. PGRN and Atsttrin prevented inflammation in multiple arthritis mouse models and inhibited TNFα-activated intracellular signaling. Collectively, these findings demonstrate that PGRN is a ligand of TNFR, an antagonist of TNFα signaling, and plays a critical role in the pathogenesis of inflammatory arthritis in mice. They also suggest new potential therapeutic interventions for various TNFα-mediated pathologies and conditions, including rheumatoid arthritis.

Progranulin (PGRN), also known as granulin epithelin precursor (GEP), PC cell–derived growth factor (PCDGF), proepithelin, and acrogranin, is an autocrine growth factor. PGRN contains seven-and-a-half repeats of a cysteine-rich motif (CX5-6CX5CCX8CCX6CCXDX2H-CCPX4CX5-6C; X, any amino acid) in the order P-G-F-B-A-C-D-E, where A-G are full repeats and P is the half-motif (*1*). PGRN is expressed in rapidly cycling epithelial cells, leukocytes, neurons (*2*), and chondrocytes (*3*). Some human cancers also express PGRN, and PGRN contributes to tumorigenesis in breast cancer, ovarian carcinoma, and multiple myeloma (*2*, *4*). PGRN plays a critical role in a variety of physiologic and disease processes, including early embryogenesis (*5*), wound healing (*6*), inflammation (*7*, *8*), host defense (*9*), and cartilage development and